

Biblioteca digitale e servizi di misurazione

Benedetta Alosi e Nunzio Femminò

*«Si misura ciò che si conosce e si valuta ciò che non si conosce;
in altre parole, la misurazione serve per conoscere gli effetti
delle decisioni organizzative prese, mentre la valutazione serve
per effettuare delle scelte, sapere quali decisioni prendere.»*

Domenico Bogliolo*

La misurazione e l'analisi dell'impatto dei servizi digitali è un tema molto dibattuto e sempre più ricorrente nella letteratura professionale recente.

L'impatto dei servizi offerti dalle Biblioteche digitali e delle risorse elettroniche messe a disposizione dell'utenza, spesso condivise con comunità più estese, quali quelle consortili, costituisce, infatti, un'attività alla quale atenei e consorzi prestano ormai un'attenzione crescente. Conoscere quanto e come vengano utilizzati servizi e risorse, e da quali comunità scientifiche e quali tipologie di utenza, significa anche avere un riscontro visibile del grado di apprezzamento delle risorse informative rese disponibili e di quanto tali risorse, così come i servizi, incontrino le necessità dei propri utenti e siano in grado di assicurare accessi rapidi e affidabili.

La rilevazione dei dati statistici che abbiamo condotto negli ultimi anni, come Centro di Ateneo per le Biblioteche dell'Università di Messina, ha riguardato prevalentemente gli accessi ad alcuni servizi on-line, implementati e gestiti dallo stesso CAB, e l'utilizzo delle risorse elettroniche acquisite dall'Ateneo in larga parte con acquisti consortili. Sin dalla sua istituzione, infatti, il CAB ha gestito centralmente l'accesso all'offerta informativa elettronica ed altre attività tipicamente connesse con la presenza di una Biblioteca Digitale. L'adesione a CIBER¹, la collaborazione tecnica con CASPUR² che ne è derivata, legata prevalentemente all'ambito delle statistiche d'uso dei periodici

* Domenico Bogliolo, *La valutazione d'efficacia e le biblioteche speciali*, in Bollettino AIB – Associazione Italiana Biblioteche, 23(1983), 1

¹ CIBER - Coordinamento Interuniversitario Basi dati & Editoria in Rete <www.uniciber.it>

² CIBER si avvale del supporto tecnico del CASPUR - Consorzio Interuniversitario Applicazioni di Supercalcolo Università e Ricerca - <www.caspur.it> che ne cura la piattaforma digitale EV (Emeroteca Virtuale) per l'accesso a tutti i periodici elettronici condivisi a livello consortile, l'archiviazione permanente delle risorse sui propri server e altri servizi, tra i quali anche l'elaborazione delle statistiche d'uso.

elettronici, e l'incontro con la comunità *Open Access* italiana³ hanno successivamente coinvolto il CAB nell'implementazione di strumenti e servizi che hanno costituito un'occasione propizia di sperimentazione sul campo di nuove tecnologie e risorse.

Nell'ambito delle attività per l'accesso aperto alla letteratura scientifica, la sperimentazione su alcuni software *open source* per la creazione di archivi aperti, e quindi la loro installazione e test, è sfociata nell'implementazione, realizzata con il sw GNU E-Prints,⁴ dell'Archivio aperto istituzionale "Messanae Universitas Studiorum" (MUS).⁵ Allo stesso tempo, il supporto tecnico offerto alle strutture dell'Ateneo nella pubblicazione di riviste digitali e ad accesso aperto ha avuto come effetto immediato il lancio degli "Atti della Accademia Peloritana dei Pericolanti – Classe di Scienze MM.FF.NN."⁶ nel nuovo formato elettronico e *open access*, e di "Communications to SIMAI Congress"⁷. La prima, rivista accademica con una sua tradizione e un prestigio da difendere da inevitabili ritardi e difficoltà legate alla pubblicazione a stampa, raccoglie le comunicazioni scientifiche dei propri soci. La seconda, in atto al primo numero nella nuova veste elettronica *open access*, contiene le comunicazioni scientifiche dei soci della Società italiana di Matematica applicata e industriale⁸.

Sempre con software *open source* è stato realizzato, infine, il portale di accesso integrato "CAB"⁹. Con la sua interfaccia unificata di interrogazione, il portale garantisce l'accesso a diverse risorse, il Catalogo elettronico collettivo, gestito con il sw *Aleph500* e

³ Le iniziative intraprese dalla Conferenza dei Rettori e, in particolare, dalla Commissione CRUI per le Biblioteche di Ateneo, hanno reso possibile l'organizzazione del Workshop nazionale "Gli Atenei per l'Open Access: verso l'accesso aperto alla letteratura di ricerca" che si è tenuto a Messina nel novembre 2004. Il Workshop e la conseguente "Dichiarazione di Messina" a sostegno dell'accesso aperto, firmata dai rettori di quasi tutte le Università italiane, hanno dato una forte spinta alla creazione di una comunità ufficiale che opera per l'affermazione delle strategie dell'accesso aperto <<http://www.aepic.it/conf/index.php?cf=1>>.

⁴ GNU E-Prints, sviluppato dalla School of Electronics and Computer Science dell'Università di Southampton (UK) <<http://www.eprints.org/software/>>.

⁵ <<http://cab.unime.it/mus>>

⁶ <<http://antonello.unime.it/atti>>

⁷ <<http://www.simai.eu>>

⁸ L'attività di supporto tecnico alla pubblicazione di riviste ad accesso aperto ha portato alla sperimentazione di alcuni sw *open source* per le riviste elettroniche, tra i quali Open Journal System (OJS <<http://pkp.sfu.ca/ocs/>>), Hyper Journal (HJ <<http://www.hjournal.org/>>), ma anche altri sw sempre *open source* per la gestione di conferenze e convegni come Open Conference System (OCS).

⁹ Il portale <<http://cab.unime.it/CDSware09>> è stato implementato con CDSware (Cern Document Server Software), sw sviluppato al Centro Europeo per la Ricerca Nucleare (CERN) di Ginevra, <<http://public.web.cern.ch/Public/Welcome.html>>. CDSware, ha tra le sue funzionalità quella di consentire l'implementazione di portali. Con CDSware il CERN ha implementato il portale "CDS" - Cern Document Server - <<http://cdsweb.cern.ch/>>. Oltre ai precedenti sw, è stato anche installato e testato per le esigenze dell'Ateneo anche Dspace <<http://www.dspace.org/>> sviluppato nel 2000 nell'ambito di un progetto congiunto MIT (Massachusetts Institute Technology) e HP (Hewlett-Packard Company)

contenente le nuove accessioni dal 2000 in poi, i cataloghi bibliografici delle Biblioteche delle Facoltà di Giurisprudenza e Scienze Politiche, non ancora importati nel catalogo elettronico, infine, i *metadati*¹⁰ dell'Archivio aperto istituzionale e di altri archivi aperti istituzionali e disciplinari, italiani e internazionali.

Una volta avviata la gestione della Biblioteca digitale, dell'OPAC e dei servizi a cui si è accennato, si è avvertita tuttavia l'esigenza di un riscontro visibile e misurabile del livello di utilizzo e di apprezzamento di quanto si stava realizzando, riscontro che si voleva estendere alle stesse home-page del Sistema Bibliotecario. Un interesse esteso era stato espresso, nel contempo, dall'intero Sistema Bibliotecario di Ateneo per la misurazione delle risorse elettroniche; la rilevazione dei dati statistici, rispondeva, infatti, all'esigenza di valutare la validità delle scelte adottate nell'acquisizione delle risorse digitali e dei criteri applicati nella distribuzione delle risorse economiche disponibili. La misurazione dell'uso delle risorse è diventata così un'attività costante e sistematica del CAB, proprio perché ritenuta in grado di fornire risposte attendibili in merito alle politiche di acquisto e di sviluppo delle collezioni elettroniche, di ripartizione di risorse tra le diverse aree disciplinari dell'Ateneo, ma anche strumento utile nella valutazione delle preferenze dei ricercatori all'interno dell'offerta informativa disponibile. I reports statistici apparivano, infatti, strumenti di rilevazioni significative sull'uso effettivo delle risorse, di chi le utilizzava, di cosa gli utenti utilizzavano e di come le utilizzavano. Rilevazioni alle quali, nelle intenzioni, e come già efficacemente sperimentato in analoghe esperienze di misurazione dell'impatto dei servizi digitali, si è pensato di affiancare altri metodi di indagine; questionari ed interviste, uniti alla rilevazione statistica, costituiscono infatti, una solida base per una più funzionale organizzazione dell'offerta informativa elettronica e dei servizi ad essa connessi.

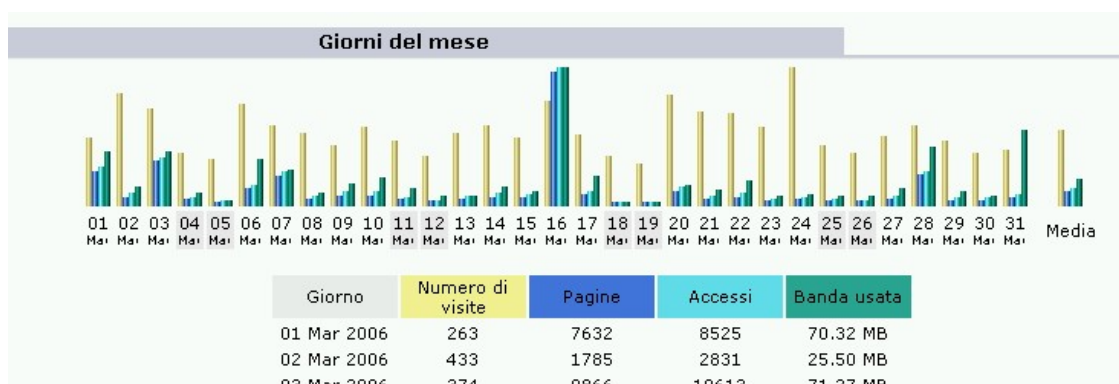
Misurazione accessi al web

All'interno dell'attività di rilevazione, la possibilità di avere uno strumento attraverso il quale condurre l'analisi dei comportamenti degli utenti della biblioteca digitale

¹⁰ "Informazioni strutturate e scandite in campi, relative ad un documento di cui permettono una più efficiente organizzazione e recupero", Wikipedia, <<http://it.wikipedia.org/wiki/Metadati>>

è sembrato un dato tutt'altro che trascurabile. In quest'ottica, è apparso così importante rintracciare un modo per analizzare gli accessi e i comportamenti degli utenti del Web, e quindi degli utenti che accedono alle pagine web del nostro sito, indipendentemente dall'utilizzo di particolari e specifici servizi. Questo tipo di rilevazione si basa sul fatto che il *navigatore* del Web lascia sempre traccia di sé, traccia che viene catturata dal file di log del web server¹¹. Di fatto, i log files ci dicono con certezza quando, come, cosa e da dove l'utente ha utilizzato un servizio.

Nell'ampia gamma di strumenti di analisi e misurazione degli accessi al web server, che vanno dai semplici contatori a strumenti molto più sofisticati e personalizzabili, la scelta è caduta su AWStats¹², software *open source*, semplice e facilmente configurabile, che analizza e misura gli accessi anche attraverso l'elaborazione di grafici.



AWStats ci ha dunque offerto una molteplicità di risposte in relazione all'uso dei servizi implementati dal CAB attraverso i propri server. Solo a titolo esemplificativo, AWStats ci ha consentito di rilevare che nel mese di marzo 2006, il Portale integrato CDSware aveva avuto 1.774 accessi con 74.692 ricerche effettuate; l'Archivio aperto istituzionale era stato visitato 470 volte; la sezione dei contributi ad alcuni convegni, che il CAB aveva gestito con il sw OCS¹³, aveva avuto 1.442 visite.

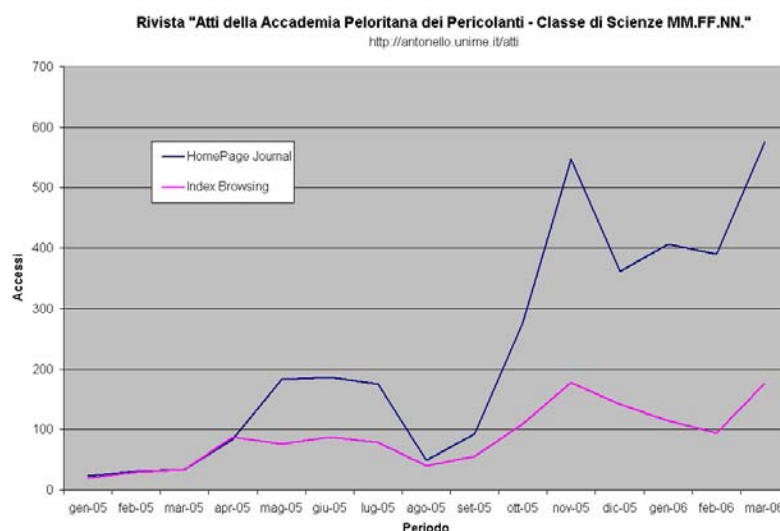
Successive e sistematiche rilevazioni ci hanno poi consentito di elaborare diverse serie stocastiche in grado di fornire una chiara visualizzazione dell'andamento degli accessi ai servizi.

¹¹ Il file di log è uno speciale archivio che tiene memoria di tutte le interazioni effettuate. Un web server è un software (e, per estensione, il computer) che si occupa di fornire, su richiesta, una pagina web

¹² <<http://awstats.sourceforge.net/>>

¹³ OCS - Open Conference System <<http://pkp.sfu.ca/ocs/>>

Analoga rilevazione è stata fatta anche sugli accessi alla rivista “Atti dell’Accademia”, per la quale vi era un alto interesse nel valutare l’impatto della sua nuova veste digitale e soprattutto della sua disponibilità ad accesso aperto. L’andamento crescente delle visite alla rivista, rilevato con AWStats, è reso palesemente visibile nel grafico riportato di seguito in cui vengono evidenziati gli accessi alla home-page e all’indice degli articoli. I dati di accesso ci hanno dato modo di dimostrare l’efficacia della scelta operata dalla Classe di Scienze dell’Accademia e dal suo Comitato editoriale, scelta tanto efficace da indurre altre classi, al momento la classe di Scienze Medico-Biologiche, a seguirne il



percorso tracciato.

Rilevazione delle statistiche d'uso delle risorse elettroniche remote (RER)

La rilevazione dei dati statistici di utilizzo delle riviste in formato elettronico, presenti nei pacchetti editoriali acquisiti dall’Ateneo, ha comportato, rispetto alle precedenti rilevazioni, un lavoro molto più complesso ma tuttavia significativo dal punto di vista dei risultati ottenuti.

Ambito di attività che sta attirando grande interesse e attenzione a livello internazionale da parte delle istituzioni accademiche e dei consorzi interuniversitari, la misurazione delle statistiche d’uso è tuttora oggetto di un numero crescente di studi, e l’evoluzione costante che la sta segnando ha prodotto recentemente iniziative importanti da

cui sono scaturiti progetti innovativi come COUNTER¹⁴ e SUSHI¹⁵ e software, come JURO¹⁶ e ScholarlyStats¹⁷.

Nonostante l'impegno della comunità accademica, la rilevazione dei dati statistici presenta tuttavia alcuni problemi ancora aperti. Trai più rilevanti: l'adesione numerosa, ma ancora non totale, degli editori al progetto COUNTER; l'esigenza, non ancora interamente ottemperata nei reports degli editori, di adeguare i formati dei files di esportazione; infine, il problema del doppio accesso, problema tipico dei consorzi che ospitano localmente le risorse acquisite e forniscono una piattaforma per l'accesso integrato ai periodici elettronici condivisi a livello consortile, in alternativa alla piattaforma disponibile sui siti degli editori.

E' questo il caso di *Emeroteca Virtuale* di CIBER e di *Cilea Digital Library*.¹⁸ Nel caso di CIBER, in particolare, l'evidente complessità della gestione dei dati statistici d'uso delle risorse consortili e i problemi di aggregazione dei dati dovuti al doppio accesso, ha indotto CIBER e CASPUR a lavorare fattivamente all'ipotesi di implementare una procedura interattiva per il rilevamento dei dati. E' così nata l'idea di creare un portale dinamico di misurazione delle statistiche d'uso di risorse elettroniche remote alla cui realizzazione hanno lavorato in collaborazione il CAB di Messina, come componente CIBER, e CASPUR.

L'attività sperimentale, che ha preceduto l'implementazione del portale, si colloca quindi all'interno di questo contesto di collaborazione che ha condotto alla realizzazione di uno strumento di elaborazione di dati statistici che da tempo la comunità CIBER avvertiva come indispensabile.

Il gruppo di lavoro statistiche del CIBER si è dunque posto come obiettivo immediato l'adeguamento dei reports statistici elaborati da CASPUR agli standards

¹⁴ COUNTER – Counting On-line of NeTworked Electronic Resources è un'iniziativa nata per fornire a bibliotecari, editori e intermediari uno strumento per facilitare la registrazione e lo scambio dei dati statistici. <<http://www.projectcounter.org/index.html>>

¹⁵ SUSHI – Standardized Usage Statistics Harvesting Initiative – standard promosso dal NISO (National Information Standard Organization) – è un modello automatico di harvesting dei dati statistici in formato COUNTER JR1 desunti dai diversi siti degli editori <http://www.niso.org/committees/SUSHI/SUSHI_comm.html>

¹⁶ JURO – Journal Usage Report On-line – è un sw open source che raccoglie manualmente i dati statistici COUNTER e ne consente la loro elaborazione. <<http://library.ust.hk/software/juro.html>>

¹⁷ ScholarlyStats è un sw commerciale sviluppato da MPS Technologies che consente l'accesso ai reports statistici dei diversi editori attraverso un'unica interfaccia. <<http://www.mpstechnologies.com/scholarlystats.htm>>

¹⁸ La piattaforma di accesso alle risorse consortili *Emeroteca Virtuale* si trova alla URL: <http://periodici.caspur.it>. La piattaforma *Cilea Digital Library* è disponibile alla URL: <<http://cdl.cilea.it/>>

internazionali attuali, e a COUNTER in particolare, che nel frattempo si è imposto come standard *de facto*, e l'implementazione di una procedura interattiva per il reperimento dei dati. In quest'ottica, la realizzazione del portale è stata preceduta dal lavoro di *parsing* dei log files dei servers di *Emeroteca Virtuale*, effettuato dal Settore automazione biblioteche di CASPUR, che ha dato come esito la produzione di reports interamente compatibili con lo standard COUNTER (R2), diversamente da quanto prodotto con il sistema SAS utilizzato sino ad allora. Il nuovo sistema di produzione semi-automatica dei reports avviene attraverso il lancio *batch* di una procedura che provvede al *parsing* e al successivo deposito dei risultati in un'area riservata, situata all'interno di una CIBER-repository, spazio accessibile via FTP offerto da CASPUR agli utenti CIBER. Dalla CIBER-repository ogni sede potrà così successivamente recuperare i propri COUNTER Journal Report 1.¹⁹

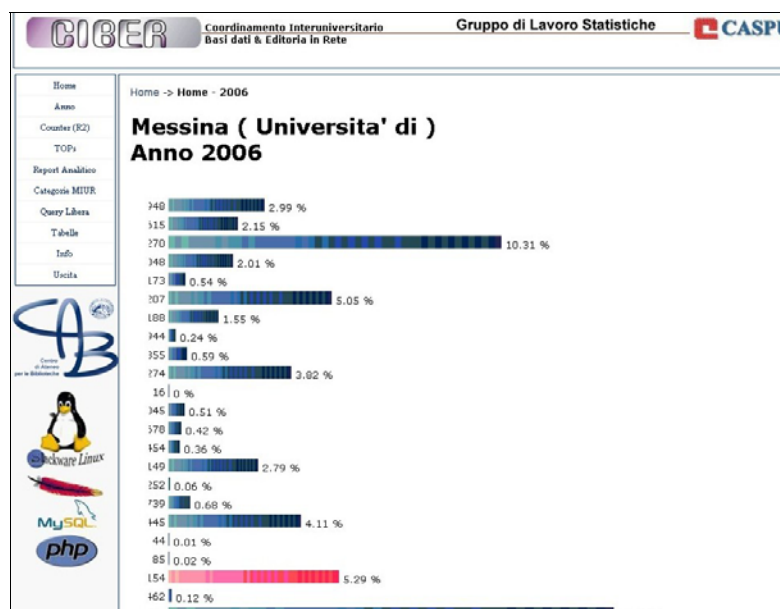
Il Portale CIBER-Stats

Per quanto semplici nell'idea e nell'utilizzo, le varie fasi di collegamento all'area FTP, il download del report, la successiva importazione ed elaborazione del file CSV,²⁰ risultavano per gli utenti ancora troppo laboriose. Inoltre, lavorando sui log files, si era anche reso evidente che la quantità delle informazioni che contenevano erano di gran lunga superiori a quelle richieste per la realizzazione dei COUNTER Journal Reports. Infatti, oltre alle indicazioni sulla data, sull'ora e sulla rivista dalla quale era stato effettuato il download, i log files fornivano anche dati importanti relativi a chi aveva scaricato l'articolo e, soprattutto, a "cosa" aveva scaricato. In definitiva, c'erano gli elementi per realizzare una "statistica del cosa" e non solo una "statistica del quanto" fornita da COUNTER. L'ipotesi di lavoro è apparsa immediatamente degna di interesse. La realizzazione, da parte di CASPUR, di una tabella riepilogativa dei *metadati* degli articoli ospitati in *Emeroteca Virtuale* e completi dell'anno di pubblicazione dell'articolo, l'ha resa poi concretamente percorribile, ancora una volta facendo ricorso a tecnologie *open source*.

¹⁹ Il Journal Report 1, Number of Successful Full-Text Article Requests by Month and Journal, è il più importante report previsto da COUNTER per la visualizzazione degli scarichi effettuati dalle riviste elettroniche. Il Journal Report 2 si riferisce invece ai tentativi di scarico, non riusciti, effettuati dagli utenti dalle riviste elettroniche non sottoscritte.

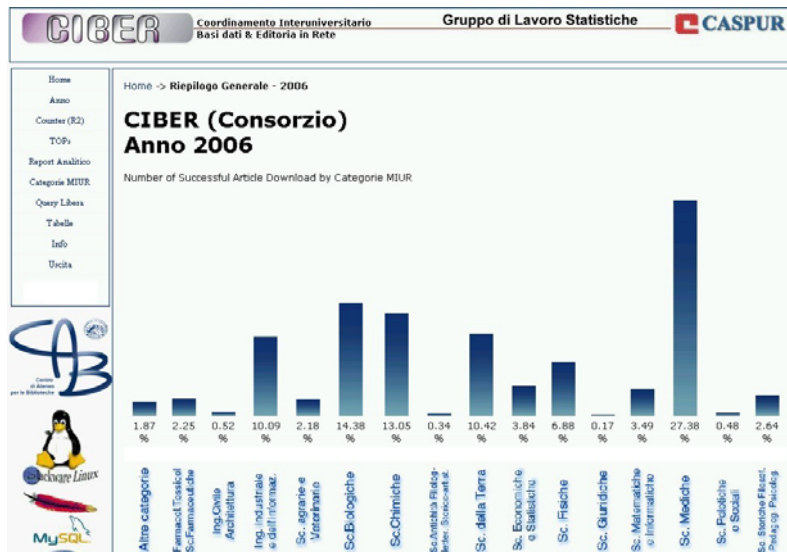
²⁰ CSV, Comma Separated Values, un particolare tipo di file testuale che prevede la virgola come separatore dei valori (campi).

L'accesso al portale è riservato agli Enti CIBER e prevede una gestione multilivello, per singolo ente e consortile. La gestione consortile aggrega i dati riguardanti gli scarichi di tutti gli enti del consorzio CIBER, funzionalità che ha semplificato notevolmente procedure sino ad allora decisamente complesse. L'immagine che segue riproduce la Home Page del portale dell'ente Messina per l'anno 2006. L'istogramma visualizzato rappresenta, in modo anonimo, la distribuzione dei downloads effettuati da tutti gli enti CIBER; in rosso è evidenziato l'ente che ha effettuato il *login*.



L'inserimento di ulteriori elementi di valutazione "esterni", come le classi disciplinari MIUR e la soggettazione della Library of Congress, recuperata, con un laborioso lavoro di dettaglio²¹ del CAB di Messina su tutti i periodici presenti in Emeroteca Virtuale, ha reso il portale estremamente flessibile e in grado di rispondere a molteplici esigenze. Ad un'esigenza di duttilità è legata, in particolare, l'implementazione della sezione "query libera", sviluppata per consentire l'interrogazione del *database* secondo parametri personalizzati. La soggettazione dei periodici ha invece consentito di valutare l'impatto delle risorse elettroniche nelle diverse comunità scientifiche, offrendo anche dati e strumenti per porre a confronto l'offerta informativa disponibile e il suo effettivo utilizzo all'interno di ogni singola area disciplinare. La *cover date* degli articoli ha permesso, infine, di individuare e misurare gli scarichi dei *backfiles* rispetto all'anno di riferimento.

²¹ Il recupero delle voci di soggetto è stato effettuato utilizzando la banca dati USAS - Ulrich's Serials Analysis System <<http://www.ulrichsweb.com/ulrichsweb/analysis/default.asp?navPage=4&>>



Va detto, infine, che il lavoro sin qui realizzato costituisce solo un tassello di una attività di sperimentazione che a livello internazionale ha prodotto risultati fortemente innovativi che hanno individuato soluzioni concrete ai problemi di integrazione dei reports provenienti da fonti diverse per gli stessi editori. JURO e SUSHI sono scaturiti proprio dall'esigenza comune di superare tali difficoltà e rappresentano, al momento, i due più importanti strumenti di riferimento.

Con l'adesione al progetto SUSHI, che comporterà l'adeguamento dei reports CIBER allo schema XML, e l'installazione di JURO, tuttora in corso di personalizzazione, il CAB di Messina si è dotato di strumenti efficaci con i quali portare avanti le attività di sperimentazione volte a realizzare una integrazione dei dati statistici quanto più automatizzata. Tali attività, rivolte e testate in prima istanza in base alle esigenze dell'ateneo, potranno avere una ricaduta più ampia e costituire un utile servizio per tutta la comunità CIBER.